A frequency-domain adaptive filter (FDAF) prediction error method (PEM) framework for double-talk-robust acoustic echo cancellation‖

Jose M. Gil-Cacho*, Toon van Waterschoot‡, Marc Moonen§, and Søren Holdt Jensen¶
*Corresponding author. KU Leuven Department of Electrical Engineering (ESAT) STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, Tel. +32 16321856, Fax +32 16321970, E-mail pepegilcacholorenzo@gmail.com
‡ KU Leuven Department of Electrical Engineering (ESAT) STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, Tel. +32 16321927, Fax +32 16321788, E-mail toon.vanwaterschoot@esat.kuleuven.be
§ KU Leuven Department of Electrical Engineering (ESAT) STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, Tel. +32 16321060, Fax +32 16 321970, E-mail marc.moonen@esat.kuleuven.be
¶ Department of Electronic Systems Aalborg University, Fredrik Bajers Vej 7, DK-9220 Aalborg, Denmark, Tel. +45 9940 8654, E-mail shj@es.aau.dk
‖EDICS: AUD-ECHO

***Abstract*— In this paper, we propose a new framework to tackle the double-talk (DT) problem in acoustic echo cancellation (AEC). It is based on a frequency-domain adaptive filter (FDAF) implementation of the so-called prediction error method adaptive filtering using row operations (PEM-AFROW) leading to the FDAF-PEM-AFROW algorithm. We show that FDAF-PEM-AFROW is by construction related to the best linear unbiased estimate (BLUE) of the echo path. We depart from this framework to show an improvement in performance with respect to other adaptive filters minimizing the BLUE criterion, namely the PEM-AFROW and the FDAF-NLMS with near-end signal normalization. One of the contributions is to propose the instantaneous pseudo-correlation (IPC) measure between the near-end signal and the loudspeaker signal. The IPC measure serves as an indication of the effect of a DT situation occurring during adaptation. We motivate the choice of FDAF-PEM-AFROW over PEM-AFROW and FDAF-NLMS with near-end signal normalization, based on performance, computational complexity and related IPC measure values. Moreover, we use the FDAF-PEM-AFROW framework to improve several state-of-the-art variable step-size (VSS) and variable regularization (VR) algorithms. The FDAF-PEM-AFROW versions significantly outperform the original versions in every simulation. In terms of computational complexity, the FDAF-PEM-AFROW versions are themselves about two orders of magnitude cheaper than the original versions.**

 This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven Research Council CoE PFV/10/002 (OPTEC), Concerted Research Action GOA-MaNet, the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P7/19 'Dynamical systems control and optimization' (DYSCO) 2012-2017, Research Project FWO nr. G.0763.12 'Wireless Acoustic Sensor Networks for Extended Auditory Communication', EC-FP7-PEOPLE Marie Curie Initial Training Network 'Dereverberation and Reverberation of Audio, Music, and Speech (DREAMS)', funded by the European Commission under Grant Agreement no. 316969, EC-FP6 project 'Core Signal Processing Training Program' (SIGNAL) and was supported by a Postdoctoral Fellowship of the Research Foundation Flanders (FWO-Vlaanderen, T. van Waterschoot). The scientific responsibility is assumed by its authors.

***Index Terms*— Acoustic echo cancellation, double-talk-robust acoustic echo cancellation, double-talk, frequency-domain adaptive filters, prediction error method, variable step size.**

I. Introduction

ACOUSTIC echo cancellation (AEC) is used in many speech communication applications where the existence of echoes degrades the speech intelligibility and listening comfort [1]. These applications range from mobile and hands-free telephony to teleconferencing and voice over IP (VoIP), and are often integrated in smartphones, tablets, notebooks, laptops, etc. The typical set-up for an acoustic echo canceler is depicted in Figure 1.

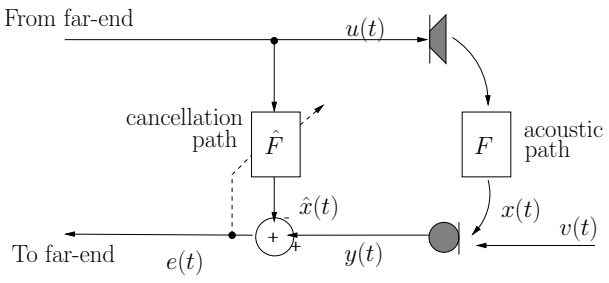

Fig. 1. Typical set-up for AEC.

A far-end speech signal $u(t)$ is played back in an enclosure (i.e., the room) through a loudspeaker. In the room there is a microphone to record a near-end speech signal, which is to be transmitted to the far-end side. An acoustic echo path between the loudspeaker and the microphone exists so that the microphone signal $y(t)$ contains an undesired echo signal $x(t)$ plus the near-end signal $v(t)$, i.e., $y(t) = x(t) + v(t)$. The echo signal $x(t)$ can be considered as the far-end speech

or loudspeaker signal $u(t)$ filtered by the echo path. An acoustic echo canceler seeks to cancel the echo signal component $x(t)$ in the microphone signal $y(t)$, ideally leading to an *echo-free* error signal $e(t)$, which is then transmitted to the far-end side. This is done by subtracting an estimate of the echo signal $\hat{x}(t)$ from the microphone signal, i.e., $e(t) = y(t) - \hat{x}(t)$. Standard approaches to AEC rely on the assumption that the echo path can be modeled by a linear FIR filter [2]-[4]. The coefficients of the echo path are collected in the parameter vector $\mathbf{f}(t) = [f_0(t), f_1(t), , ..., f_{N-1}(t)]^T \in \mathbb{R}^N$ such that $x(t) = \mathbf{f}^T(t)\mathbf{u}(t) = F(q,t)u(t)$ where $\mathbf{u}(t) = [u(t), u(t-1), \ldots, u(t-N+1)]^T$ and $F(q,t) = f_0(t) + f_1(t)q^{-1} + ... + f_{N-1}(t)q^{N-1}$ with q^{-1} the unit delay operator, i.e., $q^{-1}u(t) = u(t-1)$. An adaptive filter of sufficient order is used to provide an estimate $\hat{\mathbf{f}}(t) = [\hat{f}_0(t), \hat{f}_1(t), ..., \hat{f}_{N-1}(t)]^T \in \mathbb{R}^N$ of $\mathbf{f}$, such that the echo signal estimate is $\hat{x}(t) = \hat{\mathbf{f}}^T(t)\mathbf{u}(t) = \hat{F}(q,t)u(t)$ with $\hat{F}(q,t) = \hat{f}_0(t) + \hat{f}_1(t)q^{-1} + ... + \hat{f}_{N-1}(t)q^{N-1}$.

Practical AEC implementations often rely on computationally simple time-domain stochastic gradient algorithms, such as the least mean squares (LMS) algorithm [5], the normalized LMS (NLMS) algorithm [3], or affine projection algorithm (APA) [12], which are very sensitive to the presence of a near-end signal [2]. In general, the presence of a near-end signal, in a so-called double-talk (DT) scenario, makes the AEC adaptive filter converge slowly or even diverge. There are several approaches to tackle the problem of DT in AEC. We give here a brief explanation of five different approaches: two that are based on a *variable step size* (VSS), one based on a *variable regularization* (VR) and two based on a *best linear unbiased estimate* (BLUE) of the echo path.

The first approach is based on the so-called gradient-based VSS algorithms [6]-[11]. From this class of algorithms, the only one specifically designed for DT-robust AEC is the *projection-correlation* VSS (PCVSS), which has been proposed in [11]. PCVSS is based on the APA, and so will be referred to as PCVSS-APA. In PCVSS-APA, the adaptation rate is controlled by a measure of the correlation between instantaneous and long-term averages of the so-called *projection vectors*, i.e., gradient vectors in APA.

The second approach is based on the *non-parametric* VSS (NPVSS) algorithm proposed in [13]. Different NPVSS-based algorithms have been developed and applied for DT-robust AEC when updated with the NLMS algorithm, e.g., [14] and [15]. To increase their convergence speed, an APA version of these algorithms has been proposed, resulting in the *practical* VSS affine projection algorithm (PVSS-APA) [16].

The third approach is based on equipping the adaptive filter with a VR. Several VR algorithms have been proposed in the literature based on the derivation of an optimal regularization parameter, which sometimes need quantities or models that are difficult to obtain in practice [17], [18]. Practical implementations of these algorithms have also been proposed that employ more easily measurable quantities. One example is the APA-based VR (VR-APA) algorithm that has been proposed in [19], which incorporates the statistics of the noise into the design of the VR.

The fourth approach is based on a minimum-variance echo path estimate, i.e., the BLUE [20]. This estimate depends on the near-end signal characteristics, which are in practice unknown and time-varying [2], [21]. One approach to achieve the BLUE is based on the prediction error method (PEM) [22] for jointly estimating the echo path model and an auto-regressive (AR) model of the near-end signal. The algorithms in [2], [21] aim to whiten the near-end signal component in the microphone signal by using adaptive decorrelation prefilters that are estimated concurrently with the echo path. Among the PEM-based algorithms, the PEM adaptive filtering using row operations (PEM-AFROW) [23] is particularly interesting and it will be explained further on. Other algorithms, although not applied to DT-robust AEC, have been proposed for recursively minimizing the BLUE criterion. In fact, in [24], [25], a frequency-domain adaptive filtering (FDAF) algorithm has been obtained by first minimizing the BLUE criterion using a time-domain block stochastic gradient algorithm and then switching to the frequency domain to reduce the computational complexity. One advantage of frequency-domain adaptive filtering compared to time-domain adaptive filtering is that the step size can be normalized independently for each frequency bin. Including such a normalization in the adaptive filter update equation results in a more uniform convergence over the entire frequency range. In the sequel, the standard frequency-domain adaptive filtering [34] is referred to as FDAF-NLMS, i.e., an FDAF with a loudspeaker signal normalization factor. The algorithm proposed in [24] is therefore referred to as FDAF-NLMS *with near-end signal normalization*. The PEM-based and FDAF-based approach to achieve the BLUE will be explained further on.

One particular assumption in all these algorithms, and in most AEC applications in general, is that the near-end signal is uncorrelated with the loudspeaker signal. This assumption can truly be exploited only for infinitely long observations of ergodic and stationary processes [3]. In real AEC applications, however, the near-end

signal as well as the loudspeaker signal is a speech signal that is highly colored and non-stationary. Even when the near-end signal and the loudspeaker signal are assumed to be uncorrelated, this however does not imply that the correlation between these two signals is zero within a short-time observation window [30]. Hence, one of the contributions of this paper is to derive and define the *instantaneous pseudo-correlation* (IPC) measure between the near-end signal $v(t)$ and the loudspeaker signal $u(t)$. The IPC measure serves as an indication of the effect of a DT situation occurring during adaptation, as it will be explained.

The aim of this paper is to introduce a new framework for DT-robust AEC, which is based on an FDAF implementation of the PEM-AFROW (FDAF-PEM-AFROW). We depart from this framework to show an improvement in performance with respect to PEM-AFROW and FDAF with near-end signal normalization. Although these three algorithms are related to the BLUE, we show that FDAF-PEM-AFROW is the preferred choice for DT-robust AEC. We motivate the choice of FDAF-PEM-AFROW over PEM-AFROW and FDAF-NLMS with near-end signal normalization, based on performance improvement, computational complexity reduction and lower IPC measure values. Moreover, we use the FDAF-PEM-AFROW framework to improve the previously introduced VR-APA, PVSS-APA and PCVSS-APA leading to the VR-FDAF-PEM-AFROW, PCVSS-FDAF-PEM-AFROW and PCVSS-FDAF-PEM-AFROW respectively. The FDAF-PEM-AFROW versions significantly reduce the computationally complexity and improve the performance of the original versions in every simulation.

The paper is organized as follows. In Section II, the BLUE is explained in two subsections: in Section II-A, the basic linear regression model is reviewed, including unbiased constraints that are assumed in the derivation of typical algorithms for AEC. In Section II-B, the generalized least squares model is reviewed, which provides a framework to explain other related estimators. In Section III-A, the BLUE achieved using the PEM is considered and in Section III-B, the BLUE achieved using the FDAF-NLMS with near-end signal normalization is considered. In Section IV, the IPC measure is derived and defined, and several simulation results are shown evaluating the IPC measure in the NLMS and the FDAF-NLMS. The recursions that are used to compute the IPC measure are given in Appendix A. In Section V, the proposed FDAF-PEM-AFROW is derived. In Section VI, computer simulations are provided. In Section VI-A, we motivate the choice of the FDAF-PEM-AFROW algorithm over PEM-AFROW and FDAF-NLMS with near-end signal normalization, based on performance improvement, computational complexity reduction and lower IPC measure values. In Section VI-B more detail is provided about the selected state-of-the-art VSS and VR algorithms as well as an explanation of their FDAF-PEM-AFROW versions. Simulation results are provided with a complexity analysis comparing the following algorithms: PVSS-APA, PCVSS-APA, VR-APA, PVSS-FDAF-PEM-AFROW, PCVSS-FDAF-PEM-AFROW and VR-FDAF-PEM-AFROW. Finally, Section VII concludes the paper.

II. Best linear unbiased estimate

A. Linear unbiased estimator

We first assume that the echo path is time-invariant, $\mathbf{f}(t) = \mathbf{f}, \forall t$ over the observation window $t = 1, 2, \ldots, L$ with $L >> N$. A regression data model may be constructed as

$$\mathbf{y} = \mathbf{X}\mathbf{f} + \mathbf{v} \tag{1}$$

where

$$\mathbf{y} = [y(1), y(2), ..., y(L))]^T \tag{2}$$

$$\mathbf{v} = [v(1), v(2), ..., v(L))]^T \tag{3}$$

$$\mathbf{X} = [\mathbf{u}(1), \mathbf{u}(2), ..., \mathbf{u}(L)]^T, \tag{4}$$

with $(\cdot)^T$ the transpose operator. The near-end signal and the loudspeaker signal are usually assumed to be uncorrelated so that $\mathcal{E}\{\mathbf{X}^T\mathbf{v}\} = \mathbf{0}_{N\times 1}$, where $\mathcal{E}\{\cdot\}$ is the expected value operator. Any *linear* estimate of parameter vector $\mathbf{f}$ can be written as a linear function of the data vector $\mathbf{y}$, i.e.,

$$\hat{\mathbf{f}} = \mathbf{D}^T\mathbf{y}. \tag{5}$$

For this estimate to be *unbiased*, the $L \times N$ matrix $\mathbf{D}$ should be subjected to two constraints,

$$\mathbf{D}^T\mathbf{X} = \mathbf{I}_N \tag{6}$$

$$\mathcal{E}\{\mathbf{D}^T\mathbf{v}\} = \mathbf{0}_{N\times 1} \tag{7}$$

These constraints are often implicitly assumed to hold in most of the existing AEC algorithms.

B. Generalized least squares and BLUE

Let us now consider a *generalized least squares* (GLS) estimator given as

$$\hat{\mathbf{f}}_{\text{GLS}} = \left[\mathbf{X}^T\mathbf{M}^{-1}\mathbf{X}\right]^{-1}\mathbf{X}^T\mathbf{M}^{-1}\mathbf{y} \tag{8}$$

This estimator is quite general since the structure of the weighting matrix $\mathbf{M}$ can result in different types

of estimators [27], i.e., biased or unbiased, minimum-variance or not. As we are interested in an $\mathbf{M}$ that makes (8) the BLUE [2], we consider

$$\mathbf{M}_{\text{BLUE}} = \mathbf{R}_v = \mathcal{E}\{\mathbf{v}\mathbf{v}^T\}, \tag{9}$$

where $\mathbf{R}_v$ is the near-end signal autocorrelation matrix with symmetric Toeplitz structure, such that

$$\hat{\mathbf{f}}_{\text{BLUE}} = \left[\mathbf{X}^T\mathbf{R}_v^{-1}\mathbf{X}\right]^{-1}\mathbf{X}^T\mathbf{R}_v^{-1}\mathbf{y}. \tag{10}$$

Note that, when the near-end signal is a white noise with variance σ_v^2, the weighting matrix $\mathbf{R}_v = \sigma_v^2\mathbf{I}_L$, where $\mathbf{I}_L$ is the $L\times L$ identity matrix. If we substitute this into (10) the ordinary least squares (OLS) estimator is obtained, i.e.,

$$\hat{\mathbf{f}}_{\text{LS}} = \left[\mathbf{X}^T\mathbf{X}\right]^{-1}\mathbf{X}^T\mathbf{y} \tag{11}$$

This means that the OLS estimator achieves the BLUE only if the near-end signal is a white noise signal, otherwise the OLS estimate is suboptimal. Most of the existing AEC algorithms are based on the OLS estimator, which is assumed to be unbiased following (6) and (7), but not necessarily minimum-variance.

III. The BLUE in adaptive filtering algorithms

It has been shown [21] that the BLUE can be considered an optimal acoustic echo path estimate during DT. However, calculating the BLUE in an AEC framework is problematic due to its dependence on the near-end signal covariance matrix $\mathbf{R}_v$. Therefore, we will seek a signal transformation that diagonalizes $\mathbf{R}_v$ and include an estimation of the resulting diagonal elements in the proposed adaptive filtering algorithm. PEM-based algorithms, which have been proposed in [2], [21], provide a signal-dependent way of diagonalizing $\mathbf{R}_v$, which requires the estimation of a near-end signal model. On the other hand, FDAF-based algorithms, which have been proposed in [24], [25] are capable of diagonalizing $\mathbf{R}_v$ in a signal-independent way, after some matrix manipulation, and thus provide an attractive alternative to the PEM-based algorithms.

A. PEM-based BLUE

Note that the BLUE as in (10) usually cannot be calculated as such, because the autocorrelation matrix $\mathbf{R}_v$ is generally unknown. Therefore, $\mathbf{R}_v$ will be conveniently transformed. Let us first assume that the near-end signal is generated as $v(t) = H(q,t)w(t)$ where $w(t)$ is a white noise excitation signal with variance $\sigma_w(t)$, i.e., $\mathcal{E}\{w(t)w(t-i)\} = \delta(i)\sigma_w^2(t)$, and

$$H(q,t) = \frac{1}{A(q,t)} = \frac{1}{1 + a_1(t)q^{-1} + ... + a_{n_A}(t)q^{-n_A}} \tag{12}$$

is the near-end signal auto-regressive (AR) model, expressed as an order-n_A time-varying linear filter. The near-end signal autocorrelation matrix may then be written as

$$\mathbf{M}_{\text{PEM}} = \mathcal{E}\{\mathbf{H}(q)\mathbf{w}\mathbf{w}^T\mathbf{H}^T(q)\} \tag{13}$$

where $\mathbf{w} = [w(1), ..., w(L)]^T$ and the diagonal operator $\mathbf{H}(q)$ has the filters $H(q,t)$, $t = 1, ..., L$ on the main diagonal. In [2], [21] it has been shown that the BLUE can then be realized as

$$\hat{\mathbf{f}}_{\text{BLUE}} = \Gamma^{-1}\Psi \tag{14}$$

where

$$\Gamma = \left(\mathbf{H}^{-1}(q)\mathbf{X}\right)^T\mathbf{W}_{\text{PEM}}^{-1}\left(\mathbf{H}^{-1}(q)\mathbf{X}\right) \tag{15}$$

$$\Psi = \left(\mathbf{H}^{-1}(q)\mathbf{X}\right)^T\mathbf{W}_{\text{PEM}}^{-1}\left(\mathbf{H}^{-1}(q)\mathbf{y}\right) \tag{16}$$

with

$$\mathbf{W}_{\text{PEM}} = \begin{bmatrix} \sigma_w^2(1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_w^2(L) \end{bmatrix}, \tag{17}$$

which corresponds to a prefiltering and weighting of the tth row, $t = 1, ..., L$ of $\mathbf{X}$ and $\mathbf{y}$ with the inverse near-end signal model $H^{-1}(q,t) = A(q,t)$ and the inverse excitation signal variance $\sigma_w^{-1}(t)$. In practice, the AR model parameters in $A(q,t)$ and $\sigma_w(t)$ have to be estimated concurrently with the echo path model at each time instant $t = 1, ..., L$. The use of the PEM [22] has been proposed to achieve this, jointly providing estimates $\hat{f}(t)$, $\hat{A}(q,t)$ and $\hat{\sigma}_w(t)$. In practice, also (near-end) noise will be present in the microphone signal. Although, noise has not been included in the model for deriving the algorithm, it will be included in the simulation section.

The PEM-based approach achieving the BLUE leads to very simple time-domain stochastic gradient algorithms that feature two components (as shown in Figure 2): (1) a whitening of the near-end signal component in the microphone signal by using estimated decorrelation prefilters $\hat{A}(q,t)$, (2) a single weighting scalar in the denominator of the adaptive filter update equation using the estimated excitation signal variance $\hat{\sigma}_w(t)$ [2], [21].

In [21], it has been shown that PEM-based algorithms that achieve the BLUE, are possible if three conditions are fulfilled, (1) the near-end signal $v(t)$ can at each time instant be modeled as an AR process of order n_A, (2) the prefilter $\hat{A}(q,t)$ contains at each time instant the

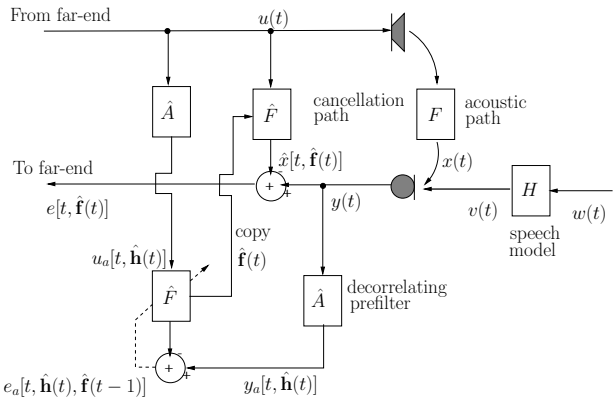

Fig. 2. Typical set-up for AEC using a decorrelation prefilter.

true AR coefficients, i.e., $\hat{A}(q,t) = A(q,t)$, and (3) the weighting scalar in the denominator of the adaptive filter update equation $\hat{\sigma}_w(t)$ is at each time instant equal to the true variance of the near-end excitation signal $w(t)$, i.e., $\hat{\sigma}_w(t) = \sigma_w(t)$. However, it seems that some of the previous conditions are difficult to fulfill in practice. In this case, the complete whitening of the near-end signal will not be achieved (i.e., $\mathbf{W}_{\text{PEM}}$ will not be diagonal) and, therefore, the use of a single weighting scalar $\hat{\sigma}_w(t)$ in the denominator of the adaptive filter update equation will not be possible. Even if the near-end signal autocorrelation matrix $\mathbf{R}_v$ could be estimated as such, its direct inversion, as needed in a typical time-domain stochastic gradient algorithm, would be prohibitive in terms of computational complexity.

B. FDAF-based BLUE

In this section, we explain how in [24] an FDAF-based algorithm has been obtained by first minimizing the BLUE criterion using a time-domain block stochastic gradient algorithm and then switching to the frequency domain to reduce the computational complexity. To this end, consider the cost function

$$J_{\text{BLUE}}(\mathbf{f}) = (\mathbf{X}\mathbf{f} - \mathbf{y})^T \mathbf{R}_v^{-1} (\mathbf{X}\mathbf{f} - \mathbf{y}), \tag{18}$$

from which the BLUE in (10) is actually obtained. Following [24] and [25], a time-domain block stochastic gradient algorithm minimizing (18) can be derived. In [24]-[26] and [29] it has been shown how FDAF-based algorithms can then be derived by rewriting a time-domain block stochastic gradient algorithm in a way that Toeplitz and circulant matrices are explicitly shown. It is important to recognize that the Toeplitz property of $\mathbf{R}_v$ is a direct consequence of the assumption that the vector $\mathbf{v}$ is wide-sense stationary [3].

Similar to [26], [29], in [24], [25] a Toeplitz matrix is diagonalized in two steps: (1) a Toeplitz matrix is transformed into a circulant matrix and (2) a circulant matrix is transformed into a diagonal matrix using the DFT. In the resulting diagonal matrix, the different diagonal elements, correspond to the near-end signal variance in different frequency bins. In FDAF-based algorithms, the near-end signal variance in each frequency bin can be straightforwardly incorporated as a normalization factor in the adaptive filter update equation. Indeed, in [24], and [25], it has been shown that such a normalization can significantly improve the performance of an FDAF-NLMS. The algorithm proposed in [24] is therefore referred to as FDAF-NLMS with near-end signal normalization. In [24] and [25], only a stationary colored near-end noise signal is considered and the near-end noise signal variance in each frequency bin is then directly estimated from the error signal $e(t)$.

IV. Instantaneous pseudo-correlation measure

The unbiased constraints in (6), (7) must be satisfied in both the PEM-based and FDAF-based BLUE. Since $\mathbf{D}$ in (7) is a function of the loudspeaker signal matrix $\mathbf{X}$, the constraint reduces to $\mathcal{E}\{\mathbf{X}^T\mathbf{v}\} = \mathbf{0}_{N\times 1}$. This means that the near-end signal should be uncorrelated with the loudspeaker signal. This assumption can truly be exploited only for infinitely long observations of ergodic and stationary processes [3]. In real AEC applications, however, the near-end signal as well as the loudspeaker signal is a speech signal that is highly colored and non-stationary. Even when the near-end signal and the loudspeaker signal are assumed to be uncorrelated, this however does not imply that the correlation between these two signals is zero within a short-time observation window [30].

It is known [32] that the correlation between these two signals causes standard adaptive filtering algorithms to converge to a *biased* solution. This means that the adaptive filter does not only predict and cancel the echo signal component in the microphone signal, but also part of the near-end signal. To analyze this, we derive and define the *instantaneous pseudo-correlation* (IPC) measure between the near-end signal $v(t)$ and the loudspeaker signal $u(t)$. It should be clear that the IPC measure is a performance measure used in simulations, hence every signal is needed separately.

To this end, we first consider the time-domain LMS algorithm update equation given as

$$e(t) = y(t) - \hat{\mathbf{f}}^T(t-1)\mathbf{u}(t) \tag{19}$$

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \mu e(t)\mathbf{u}(t). \tag{20}$$

By repeated application of (20), and assuming $\hat{f}(0) = 0$, it is straightforwardly shown that

$$\hat{\mathbf{f}}(t-1) = \mu \sum_{i=1}^{t-1} e(i)\mathbf{u}(i) \tag{21}$$

so that

$$e(t) = y(t) - \mu \sum_{i=1}^{t-1} e(i) \left(\mathbf{u}^T(i)\mathbf{u}(t)\right) \tag{22}$$

where $y(t) = x(t) + v(t)$.

Let us now consider a second scenario where the near-end signal is absent, i.e., $v(i) = 0$, $i = 1, ..., t$, and so the microphone signal is equal to the echo signal $y(i) = x(i), i = 1, ..., t$. Applying the LMS algorithm to this scenario (denoted by means of a subscript 'x') then leads to

$$e_x(t) = x(t) - \mu \sum_{i=1}^{t-1} e_x(i) \left(\mathbf{u}^T(i)\mathbf{u}(t)\right) \tag{23}$$

Here, the $e_x(t)$ corresponds to the residual echo, when an 'ideal' filter adaption is run, i.e., in the absence of a near-end signal. In a DT scenario, the $e(t)$ can then be identified as the sum of $e_x(t)$ and an extra $e_v(t)$, which is due to the DT. From (22) and (23) it follows that

$$\begin{aligned} e_v(t) &= e(t) - e_x(t) \\ &= y(t) - x(t) - \mu \sum_{i=1}^{t-1} \big(e(i) - e_x(i)\big) \big(\mathbf{u}^T(i)\mathbf{u}(t)\big) \\ &= v(t) - \mu \sum_{i=1}^{t-1} e_v(i) \big(\mathbf{u}^T(i)\mathbf{u}(t)\big), \end{aligned} \tag{24}$$

which effectively corresponds to applying the LMS algorithm to a scenario with only the near-end signal $v(t)$ and no echo signal, i.e., $x(i) = 0$ $i = 1, ..., t$.

The aim of AEC is to have the best possible echo cancellation, i.e., the smallest possible $e_x(t)$, next to preserving the near-end signal $v(t)$. Hence, ideally $e_v(t)$ should be equal to $v(t)$. From (24) it follows that this implies that ideally

$$\mu \sum_{i=1}^{t-1} e_v(i) \left(\mathbf{u}^T(i)\mathbf{u}(t)\right) = 0 \tag{25}$$

Based on (25), we define a signal

$$z(t) = \frac{\mu}{N} \sum_{i=1}^{t-1} e_v(i) \left(\mathbf{u}^T(i)\mathbf{u}(t)\right) \quad = \frac{\mu}{N} \hat{\mathbf{f}}^T(t-1)\mathbf{u}(t), \tag{26}$$

which resembles a correlation function, hence the name IPC. The signal $z(t)$ contains the factor $1/N$ in (26) to normalize the inner product with respect to the filter length. Consequently, we want $z(t)$ to be as small as possible, ideally $z(t) = 0$ as (25). Although, unfortunately, this is not the case in practice we show that several algorithms can help to reduce the signal $z(t)$.

We consider the recursions given in Appendix A where it is shown how the signal $z(t)$ is computed in the NLMS and the FDAF-NLMS based on a derivation similar to the above derivation for LMS. As an IPC measure within an adaptation loop, we consider

$$\text{IPC measure} = 10 \log_{10} \frac{\sum_{t=1}^{L} z(t)^2}{\sum_{t=1}^{L} v(t)^2} \quad \text{dB}, \tag{27}$$

i.e., the estimated variance of the signal $z(t)$ normalized w.r.t. the estimated variance of the near-end signal with L the total length of the signals. The normalization makes the IPC measure independent of the level of the near-end signal. In what follows, we show that within their adaptation loops the NLMS, and the FDAF-NLMS show different values of the IPC measure.

$N = 1001$	White noise	Colored noise	Speech
NLMS	-17.6 dB	-17 dB	-15.4 dB
FDAF-NLMS	-26.5 dB	-26 dB	-23.5 dB

TABLE I
IPC MEASURE FOR THE NLMS AND THE FDAF-NLMS. THE LOUDSPEAKER SIGNAL IS A FEMALE SPEECH SIGNAL AND THE NEAR-END SIGNAL IS A WHITE NOISE SIGNAL, A COLORED NOISE SIGNAL OR MALE SPEECH SIGNAL.

Table I shows the values of the IPC measure for the NLMS and the FDAF-NLMS. Here, the loudspeaker signal is a female speech signal and the near-end signal is either a white noise signal, a colored noise signal or male speech signal, and $N = 1001$ equal to the length of the impulse response of the system under study. The smallest IPC measure is obtained using the FDAF-NLMS in each scenario. As expected, it can also be observed that the largest IPC measure is obtained with a speech signal, which is more correlated to the male speech than to noise. In addition, the values of the IPC measure for different filter lengths are shown in Figure 3. High IPC measure means that the adaptive filter would not only predict and cancel the echo signal component in the microphone signal, but also part of the near-end signal. It is shown that by increasing the filter length, the IPC measure in NLMS and FDAF-NLMS is reduced. This reduction is clearly higher in the FDAF-NLMS, which highlights the decorrelation properties of FDAF-based algorithms and explains why they are suitable for DT-robust AEC.

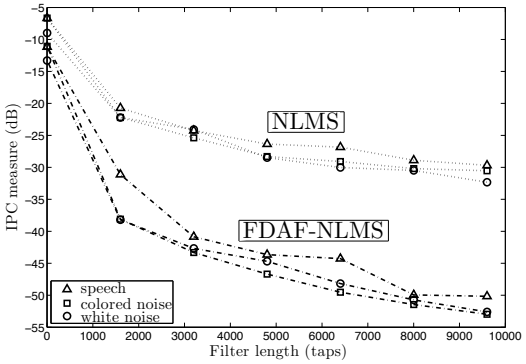

Fig. 3. IPC measure for the NLMS and the FDAF-NLMS. The loudspeaker signal is a female speech signal and the near-end signal is a white noise signal, a colored noise signal or male speech signal.

V. The FDAF-PEM-AFROW algorithm

Among the PEM-based algorithms proposed for DT-robust AEC in [2] and [21], the PEM-AFROW [23] is particularly interesting. When the near-end signal $v(t)$ is a speech signal, which is considered to be short-term stationary, the near-end signal model $A(q,t)$ does not need to be re-estimated at each time instant t. That is, instead of identifying the near-end signal model recursively, this can also be identified non-recursively on a block of loudspeaker and microphone samples. This is the idea behind the PEM-AFROW, which estimates $A(q,t)$ in a block-based manner, using a block length that approximates the stationary interval of the near-end signal. The coefficients of the prefilter $\hat{\mathbf{a}}(t) = [\hat{a}_1(t), ..., \hat{a}_{n_A}]^T$ (12) and $\hat{\sigma}_w(t)$ are efficiently computed using the Levinson-Durbin algorithm [3]. The Levinson-Durbin algorithm is a well-known procedure in linear algebra to recursively calculate the solution to an equation involving a Toeplitz matrix. The algorithm proposed in this paper extends the PEM-AFROW using the well-known FDAF with gradient constraint, based on the overlap-save method where the size of each input block is N and so the DFT size is $M = 2N$. For a complete description of the PEM-AFROW and the FDAF the reader is referred to [23] and [34], respectively. A complete description of the algorithm proposed here, which is refereed to as FDAF-PEM-AFROW, is provided as Algorithm 1, where $\mathcal{F}(\mathcal{F}^{-1})$ represents a DFT(IDFT) operation, '$\circ$' represents an element-wise multiplication, $[\cdot]_{a:b}$ represents a range of samples within a vector, k is the block index, capital letters represent frequency-domain variables, lower-case letters represent time-domain variables, boldface letters represent vector variables and non-boldface letters represent scalar variables. A vector of frequency-domain

Algorithm 1 FDAF-PEM-AFROW

1: Initialize: $M = 2N$, $\hat{\mathbf{F}}(0) = \mathbf{e}^T(0) = \mathbf{S}_{U_a}(0) = \mathbf{0}_{M\times 1}$.
2: Vectors $\mathbf{u}(k)$, $\mathbf{y}(k)$, and hence $\mathbf{u}_a(k)$ and $\mathbf{y}_a(k)$, are length-M vectors satisfying the overlap-save condition in FDAF, $\mathbf{u}(k) = [u(kN - N + 1), ...u(kN + N)]^T$, $\mathbf{y}(k) = [y(kN - N + 1), ...y(kN + N)]^T$, $\mathbf{u}_a(k) = [u_a(0), ...u_a(M-1)]^T$ and $\mathbf{y}_a(k) = [y_a(0), ...y_a(M-1)]^T$.
3: **for** $k = 1, 2, ...$ **do**
4: $\quad \mathbf{e}(k) = \left[\mathbf{y}(k) - \mathcal{F}^{-1}\left\{\mathcal{F}\{\mathbf{u}(k)\} \circ \hat{\mathbf{F}}(k-1)\right\}\right]_{N+1:M}$
5: $\quad \mathbf{r}(k) = [\mathbf{e}^T(k), \quad \mathbf{e}^T(k-1)]^T$
6: $\quad [\hat{\mathbf{a}}(k) \quad \hat{\sigma}_w(k)] = \text{Levinson-Durbin}\{[\mathbf{r}(k)]_{1:P}, n_A\}$
7: $\quad$ **for** $m = 0, ..., M-1$ **do** (Decorrelation prefilter)
8: $\quad\quad u_a(m,k) = [u(kN+1+m), ..., u(kN+1+m-n_A)]\hat{\mathbf{a}}(k)$
9: $\quad\quad y_a(m,k) = [y(kN+1+m), ..., y(kN+1+m-n_A)]\hat{\mathbf{a}}(k)$
10: $\quad$ **end for**
11: $\quad \mathbf{U_a}(k) = \mathcal{F}\{\mathbf{u}_a(k)\}$
12: $\quad \mathbf{e}_a(k) = \left[\mathbf{y}_a(k) - \mathcal{F}^{-1}\left\{\mathbf{U_a}(k) \circ \hat{\mathbf{F}}(k-1)\right\}\right]_{N+1:M}$ (Prediction-error signal)
13: $\quad \mathbf{E}_a(k) = \mathcal{F}\{\left[\mathbf{0}_N \quad \mathbf{e}_a^T(k)\right]^T\}$
14: $\quad$ **for** $m = 0, 1, ..., M-1$ **do**
15: $\quad\quad S_{U_a}(m,k) = \lambda_0 S_{U_a}(m,k-1) + (1-\lambda_0)|U_a(m,k)|^2$ (Recursive power estimate of loudspeaker signal)
16: $\quad\quad G(m,k) = (S_{U_a}(m,k) + \hat{\sigma}_w(k) + \alpha)^{-1}$ (Normalization factor)
17: $\quad$ **end for**
18: $\quad \hat{\mathbf{F}}(k) = \hat{\mathbf{F}}(k-1) +$
19: $\quad \mu_0 \mathcal{F}\left\{\left[\left(\mathcal{F}^{-1}\{\mathbf{G}(k) \circ \mathbf{U}_a{}^*(k) \circ \mathbf{E}_a(k)\}\right)_{1:N}^T \quad \mathbf{0}_N\right]^T\right\}$
20: **end for**

variables would then be $\mathbf{G}(k)$, where each entry would be $G(m,k)$, with m the frequency bin.

In *line* 5, two length-N error vectors are concatenated in $\mathbf{r}(k)$ to calculate both the order-n_A AR coefficients and the near-end excitation signal variance (i.e., $[\hat{\mathbf{a}}(k) \quad \hat{\sigma}_w(k)]$, respectively) using the Levinson-Durbin algorithm, where P is the block length used to estimate $A(q,t)$ with $N \leq P \leq 2N$. The AR model coefficients are used to prefilter the loudspeaker signal $u(t)$ and the microphone signal $y(t)$ to obtain the length-$2N$ vectors $\mathbf{u}_a$ and $\mathbf{y}_a$. With these prefiltered signals, a standard FDAF algorithm with gradient constraint is then performed in *line* 11 to *line* 19.

It is worth noticing that the normalization factor in *line* 16 contains three terms: (1) the frequency-dependent loudspeaker signal power estimate $S_{U_a}(m,k)$, (2) the non-frequency-dependent near-end excitation signal variance estimated in *line* 6 and (3) a small regularization term α to avoid division by zero. The latter term introduces some bias, which is assumed to be very small. It is noted that in FDAF-based algorithms, the near-end signal variance estimate in each frequency bin can be straightforwardly incorporated in the normalization factor of the adaptive filter update equation. In [24] and [25], the near-end noise signal variance in each frequency bin is directly estimated in the frequency domain from the error signal $e(t)$. On the other hand, the FDAF-PEM-AFROW uses the same near-end signal

variance estimate in each frequency bin, i.e., the near-end excitation signal variance, which is readily available from the Levinson-Durbin algorithm. This is because PEM-AFROW-based algorithms feature a whitening of the near-end signal by applying the prefilters $\hat{A}(q,t)$ (as shown in Figure 2). In the next section, the superior performance of the FDAF-PEM-AFROW compared to PEM-AFROW and FDAF-NLMS with near-end signal normalization is demonstrated.

VI. Simulation results

Simulations are performed using speech signals. The sampling frequency in every simulation is 8 kHz, the far-end signal is a female speech signal and the near-end signal is a male speech signal. The microphone signal consists of three concatenated segments of speech: the first and third 12.5-s segments consist of echo only, the second segment is the sum of echo and near-end signal generating a DT situation of 13 s. The 1001-taps acoustic path has been measured in a room. Notice that throughout the paper, we assume a sufficient-order condition for the acoustic path model, i.e., $N = 1001$. The tuning parameters of every algorithm are chosen to have a similar initial convergence that also leads to acceptable DT robustness. Their specific values are shown only for the FDAF-PEM-AFROW versions. The Matlab functions, and scripts with tuning parameters, that are used to generate the figures in this section are available online [1].

The *Misalignment* (MSL) is used as performance measure to evaluate the different algorithms. The MSL between the estimated echo path $\hat{\mathbf{f}}(t)$ and the true echo path $\mathbf{f}$ represents the accuracy of the echo path estimation and is defined as,

$$\text{MSL}(t) = 10\log_{10}\frac{\left\|\hat{\mathbf{f}}(t)-\mathbf{f}\right\|_2^2}{\|\mathbf{f}\|_2^2}, \tag{28}$$

where the l_2 norm $\|\mathbf{x}\|_2^2 = \sum_{i=1}^{N} x_i^2$. Moroever, the signal-to-noise ratio is defined as,

$$\text{SNR} = 10\log_{10}\frac{\|\hat{\mathbf{x}}(t)\|_2^2}{\|\mathbf{n}(t)\|_2^2}, \tag{29}$$

where $n(t)$ is computer-generated white noise that has been added to the microphone signal. The signal-to-echo ratio is defined as,

$$\text{SER} = 10\log_{10}\frac{\|\hat{\mathbf{x}}(t)\|_2^2}{\|\mathbf{v}(t)\|_2^2}. \tag{30}$$

[1] $http://homes.esat.kuleuven.be/\sim dspuser/abstract13-13\_2.html$

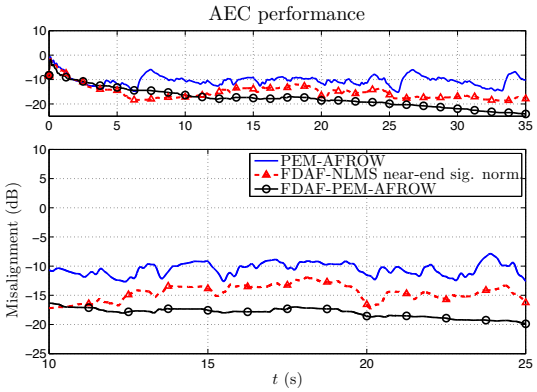

Fig. 4. AEC performance using speech signals (far-end and near-end) between 12.5 and 25 s. Bursting DT occurs at 10 dB SER and white near-end noise at 30 dB SNR, $N = 1001$, $n_A = 1$, $P = 160$. The value of SNR is typical in AEC applications and the value of SER results in a moderate level. The upper part shows the full-length simulation and the bottom part shows a zoom-in coinciding with the DT. Comparison among three algorithms constructed from the BLUE framework: PEM-AFROW, FDAF-NLMS with near-end signal normalization, and FDAF-PEM-AFROW.

A. Choice of the FDAF-PEM-AFROW algorithm

In this section, we compare the FDAF-PEM-AFROW to the FDAF-NLMS with near-end signal normalization [24] and to the PEM-AFROW [23]. The comparison is done based on both performance and complexity and it is structured to investigate whether or not PEM-AFROW and FDAF-NLMS with near-end signal normalization, could be a better option for DT-robust AEC than the proposed FDAF-PEM-AFROW algorithm. It is noted that these three algorithms are constructed from the BLUE framework.

- Performance

Figure 4 shows the MSL performance comparison between the PEM-AFROW, the FDAF-NLMS with near-end signal normalization and the FDAF-PEM-AFROW, where the upper part shows the full-length simulation and the bottom part shows a zoom-in coinciding with the DT situation. FDAF-PEM-AFROW obtains the lowest MSL values of all the other at each time instant both in single talk and in DT. In the bottom part of Figure 4, it is clearly seen that FDAF-PEM-AFROW generally outperforms, by $4-6$ dB, the FDAF-NLMS with near-end signal normalization. Compared to PEM-AFROW, FDAF-PEM-AFROW achieves a $7-9$ dB improvement. FDAF-NLMS with near-end signal normalization appears to outperform the PEM-AFROW.

- Computational complexity

Algorithm	Computation	Total
PEM-AFROW	$\left(8+\frac{4P+2n_A+1}{P}\right)N+\frac{1}{P}n_A^2+$ $\left(4+\frac{4P+2}{P}\right)n_A+\frac{P-1}{P}+10$	12049
FDAF-NLMS near-end sig. norm.	$1/N(18M\log_2 M+18M+$ $3M)$	438
FDAF-PEM-AFROW	$1/N(18M\log_2 M+18M)+$ $1/N(n_A^2+n_A(4+4M))$	483

TABLE II

COMPLEXITY COMPARISON BY THE NUMBER OF FLOPS PER RECURSION. ONE FFT/IFFT IS A $3M\log_2 M$ COMPLEXITY OPERATION. $N=1001, n_A=1$ AND $P=160$.

$N=1001, n_A=1$	White	Colored	Speech
PEM-AFROW	-30.2 dB	-30.8 dB	-30.8 dB
FDAF-PEM-AFROW	-30.8 dB	-48.0 dB	-49.8 dB
$N=1001, n_A=12$	White	Colored	Speech
PEM-AFROW	-30.2 dB	-31.4 dB	-31.1 dB
FDAF-PEM-AFROW	-34.6 dB	-46.0 dB	-50.6 dB

TABLE III

IPC MEASURE FOR PEM-BASED ALGORITHMS. THE LOUDSPEAKER SIGNAL IS A FEMALE SPEECH SIGNAL AND THE NEAR-END SIGNAL IS A WHITE NOISE SIGNAL, A COLORED NOISE SIGNAL OR MALE SPEECH SIGNAL. $N=1001, n_A=1$ AND $n_A=12$.

Table II shows the complexity analysis of the three algorithms. It is shown that, for a typical choice of the algorithm parameters, FDAF-PEM-AFROW has a similar complexity as FDAF-NLMS with near-end signal normalization, for a significant performance improvement. Moreover, FDAF-PEM-AFROW is 3.2 times cheaper than PEM-AFROW, for an even more significant performance improvement.

- IPC measure in PEM-AFROW-based algorithms

Table III shows the values of the IPC measure (27) for the PEM-AFROW and the proposed FDAF-PEM-AFROW. This is to show the impact that either the prefilter or the type of adaptation (i.e., time-domain or frequency-domain) has on the resulting IPC measure. The IPC measure is calculated without using the near-end signal variance estimate. Calculations are performed in the same scenario as before with $N=1001$ and two different AR model orders, $n_A=1$ and $n_A=12$. The recursions for computing $z(t)$ in PEM-based algorithms are given in Appendix A.

The results when the near-end signal is a white noise signal are similar when using $n_A=1$ and $n_A=12$. On the other hand, in the colored noise signal case and in the speech signal case the improvement is much higher. Increasing the near-end signal model order, from $n_A=1$ to $n_A=12$ also reduces the IPC measure in the speech signal case. The fact that a prefilter is included in the recursions seems to significantly reduce the IPC measure in the colored noise signal case and in the speech signal case. Interestingly enough, the IPC measure in the white noise signal case is similar for FDAF-NLMS in Table I and for the FDAF-PEM-AFROW in Table III.

To conclude, comparing Table I to Table III it seems that the IPC measure in the time-domain NLMS is higher than that of the time-domain PEM-AFROW. The reason appears to be that PEM-AFROW includes a prefiltering operation. However, the IPC measure in PEM-AFROW is higher than that of FDAF-NLMS. The reason is that FDAF-NLMS is implemented in the frequency domain, which seems to reduce the IPC measure as seen in Figure 3. Finally and gathering both a prefilter and a frequency-domain implementation, the FDAF-PEM-AFROW has the lowest IPC measure of all. Although the three algorithms are constructed from the BLUE framework and should therefore obtain the minimum variance echo path estimate during DT, it turns out that the differences between them are clear (as shown in Figure 4). In the PEM-AFROW, it is clear that fulfilling the three conditions to achieve the BLUE in practice is very difficult. This fact seems to affect the time-domain PEM-AFROW much more than the FDAF-PEM-AFROW. The FDAF-NLMS with near-end signal normalization performs better than PEM-AFROW as the conditions to achieve the BLUE seem to be less restricting. It seems however that the assumption of the near-end signal being wide-sense stationary and obtaining the near-end signal variance estimate per frequency bin are also difficult to fulfill in practice. FDAF-NLMS with near-end signal normalization is clearly outperformed by FDAF-PEM-AFROW, which has the benefits of being implemented in the frequency-domain, featuring a prefilter and including the non-frequency-dependent near-end signal variance estimate.

B. Results from VSS algorithms

In this section, we explain the proposed FDAF-PEM-AFROW versions of three state-of-the-art algorithms: variable regularization (VR-APA) [19], practical variable step size (PVSS-APA) [16], projection-correlation variable step size (PCVSS-APA) [11]. It is important to notice that the FDAF-PEM-AFROW given in Algorithm 1 uses the inverse of the estimated variance of the near-end excitation signal $w(t)$, i.e., $\hat{\sigma}_w^{-1}(t)$ to account for the variance in the estimation so as to obtain the BLUE of the echo path. $\hat{\sigma}_w(t)$ is estimated directly using the Levinson-Durbin algorithm and is subsequently used in the adaptation gain $G(m,k)$. On the other hand, in the PVSS-APA [16] and VR-APA [19] the near-end signal variance $\sigma_v(t)$ needs to be estimated instead.

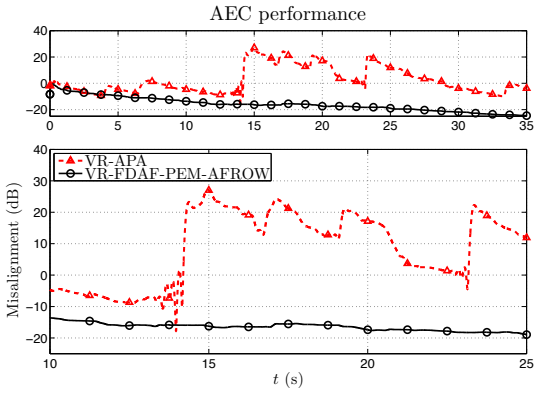

Fig. 5. AEC performance using speech signals (far-end and near-end) between 12.5 and 25 s. The original APA and the FDAF-PEM-AFROW version of variable regularization are compared when bursting DT occurs at 10 dB SER and white near-end noise at 30 dB SNR. The upper part shows the microphone signal contributions where the dark solid line is the echo and the light green is the near-end signal.

The three selected VSS algorithms are implemented using FDAF-PEM-AFROW and show different levels of improvement as compared to the original versions using APA with projection order $K = 4$. In Figure 5, the upper part shows the microphone signal contributions where the dark solid line is the echo and the light green is the near-end signal. In Figure 6(a) and Figure 6(b) the upper part shows the full-length simulation and the bottom part shows a zoom-in coinciding with the DT situation. In the references to the original versions, a deep explanation of the original algorithms, and the effect of the main parameters on the adaptation, is given. In the FDAF-PEM-AFROW versions, the most important parameter is the step size μ_0, which controls the convergence speed. The other parameters are mostly used to fine-tune the smoothness of the curve. In every simulation we use the parameters, $N = 1001$, $n_A = 1$, $P = 160$.

- VR-FDAF-PEM-AFROW (Algorithm 2):

In [19], a practical algorithm to design a VR factor for the APA has been proposed. The condition to derive the VR-APA is to minimize the difference between the estimated and true filter coefficients. For this, it is assumed that the l_2 norm of the *a posteriori* error is equal to the near-end noise signal variance, similar to the condition imposed in [13], [16]. The VR-APA performance has been compared to the performance of existing techniques [17], [18] [35], demonstrating the effectiveness of VR-APA. The implementation of the VR-FDAF-PEM-AFROW is straightforward since the estimate of the near-end excitation signal variance is calculated using the Levinson-Durbin algorithm. The VR

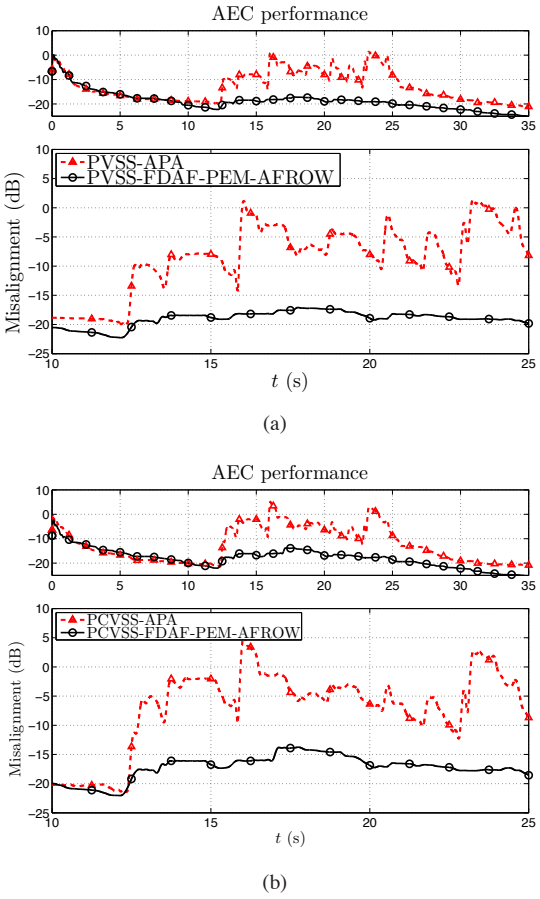

Fig. 6. AEC performance using speech signals (far-end and near-end) between 12.5 and 25 s. The original APA and the FDAF-PEM-AFROW versions are compared when bursting DT occurs at 10 dB SER and white near-end noise at 30 dB SNR. In (a) and (b) The upper part shows the full-length simulation and the bottom part shows a zoom-in coinciding with the DT. (a) Practical VSS. (b) Projection-correlation VSS.

parameter is obtained in *line* 8 and included in the normalization factor in *line* 9 as shown in Algorithm 2.

In this particular case, the performance of VR-APA, shown in Figure 5, is very poor. The algorithm is difficult to tune such that it has a comparable initial convergence curve as the VR-FDAF-PEM-AFROW. On the other hand, it is obvious that VR-FDAF-PEM-AFROW significantly outperforms VR-APA, turning the latter into an algorithm suitable for DT situations as well.

- PVSS-FDAF-PEM-AFROW (Algorithm 3):

In PVSS-APA [16], the condition that is imposed to preserve the near-end component in the echo-compensated signal is to set the *a posteriori* error equal to the near-end signal. As satisfying this condition is not possible in

practice, an estimate of the near-end signal variance is calculated and compared to an estimate of the variance of the error signal [13], [16]. The PVSS-FDAF-PEM-AFROW version of the original algorithm is given in Algorithm 3. The main steps of this algorithm are: (1) the near-end signal variance estimation, which in PVSS-FDAF-PEM-AFROW is calculated using the Levinson-Durbin algorithm, and (2) the VSS calculation in *line* 7, which is frequency-dependent in PVSS-FDAF-PEM-AFROW.

The convergence of PVSS-FDAF-PEM-AFROW, shown in Figure 6(a), is similar to the convergence for the PVSS-APA. During DT PVSS-FDAF-PEM-AFROW clearly outperforms PVSS-APA.

- PCVSS-FDAF-PEM-AFROW (Algorithm 4):

PCVSS-APA [11] belongs to the gradient-based VSS algorithms. It appears that PCVSS outperforms the algorithms given in [6] and [36] in DT situations and moreover it does not rely on any signal or system model so it is claimed to be easy to control in practice. The adaptation rate is controlled by a measure of the correlation between instantaneous and long-term averages of the so-called *projection vectors*, i.e., gradient vectors in APA. The three main features of PCVSS-FDAF-PEM-AFROW (Algorithm 4) are represented in (1) *line* 7 where the correlation of the current and past gradient estimates is calculated, (2) *line* 8 where the current step size is generated and (3) *line* 9 where the control logic to bound its value is applied.The two algorithms, shown in Figure 6(b), have a similarly fast initial convergence. During DT, the improvement of PCVSS-FDAF-PEM-AFROW w.r.t. PCVSS-APA becomes apparent. PCVSS-FDAF-PEM-AFROW outperforms PCVSS-APA.

C. Complexity analysis

The complexity analysis in Table IV shows that, for a typical choice of the algorithm parameters, the FDAF-PEM-AFROW algorithms are about two orders of magnitude cheaper than the corresponding original algorithms. Hence, it is concluded that the PVSS-FDAF-PEM-AFROW, PCVSS-FDAF-PEM-AFROW and VR-FDAF-PEM-AFROW algorithms are to be preferred over the original algorithms during DT situations. Moreover, based on FDAF-PEM-AFROW, a highly robust algorithm has been proposed in [37] using a Wiener variable step size (WVSS) and a gradient spectral variance smoothing (GSVS) for DT-robust AEC and for acoustic feedback cancellation (AFC). In their AEC simulations, the WVSS-GSVS-FDAF-PEM-AFROW algorithm has obtained robustness and smooth adaptation in highly adverse scenarios such as in bursting DT at high levels, and in a change of acoustic path during continuous DT using white as well as colored non-stationary near-end noise.

VII. Conclusion

In this paper, we have proposed a new framework to tackle the problem of double-talk (DT) in acoustic echo cancellation (AEC). It is based on a frequency domain adaptive filtering (FDAF) implementation of the so-called PEM-AFROW algorithm (FDAF-PEM-AFROW). It has been shown that the FDAF-PEM-AFROW minimizes the BLUE criterion and so provides an optimal acoustic echo path estimate during DT. The FDAF-PEM-AFROW algorithm shows an improved performance with respect to the PEM-AFROW and FDAF-NLMS with near-end signal normalization. Although these three algorithms are constructed from the BLUE framework and should therefore obtain the minimum variance echo path estimate during DT, in practice clear differences between them are observed. In PEM-AFROW, it is clear that fulfilling the three conditions to achieve the BLUE in practice is very difficult. This fact affects the PEM-AFROW much more than the FDAF-PEM-AFROW. The FDAF-NLMS with near-end signal normalization performs better than PEM-AFROW as the conditions to achieve the BLUE seem to be less restricting. However, the FDAF-NLMS with near-end signal normalization is still outperformed by FDAF-PEM-AFROW, which has the benefits of being implemented in the frequency-domain, featuring a prefilter and including the non-frequency-dependent near-end excitation signal variance estimate. We have shown that the instantaneous pseudo-correlation (IPC) measure between the near-end signal and the loudspeaker signal is significantly reduced when using the combination of FDAF and PEM, as done in FDAF-PEM-AFROW.

Finally, we have used the FDAF-PEM-AFROW framework to improve several state-of-the-art variable step-size (VSS) and variable regularization (VR) algorithms, in particular, the APA-based projection-correlation VSS (PCVSS-APA), the practical VSS affine projection algorithm (PVSS-APA) and the APA-based VR (VR-APA). It has been shown that the FDAF-PEM-AFROW algorithms significantly outperform the corresponding original algorithms in every simulation. In terms of computational complexity the FDAF-PEM-AFROW versions are themselves about two orders of magnitude cheaper than the original versions.

Algorithm 2 Variable regularization [19], using FDAF-PEM-AFROW (VR-FDAF-PEM-AFROW)

1: Initialize: $\boldsymbol{S}_{U_a} = \boldsymbol{S}_{E_a} = \mathbf{0}_{M\times 1}$.
2: **for** $k = 1, 2, ...$ **do**
3: Perform lines $4-14$ from Algorithm 1
4: $\mathbf{\Phi}(k) = {\mathbf{U}_a}^*(k) \circ \mathbf{E}_a(k)$
5: **for** $m = 0, 1, ..., M-1$ **do**
6: $S_{U_a}(m,k) = \lambda_0 S_{U_a}(m,k-1) + (1-\lambda_0)|U_a(m,k)|^2$
7: $S_{E_a}(m,k) = \lambda_1 S_{E_a}(m,k) + (1-\lambda_1)\,|E_a(m,k)|^2$
8: $\alpha(m,k) = S_{U_a}(m,k)\dfrac{\sqrt{\hat{\sigma}_w(k)}}{\sqrt{S_{E_a}(m,k)} - \sqrt{\hat{\sigma}_w(k)}}$
9: $G(m,k) = \left(S_{U_a}(m,k) + \alpha(m,k)\right)^{-1}$
10: **end for**
11: $\hat{\mathbf{F}}(k) = \hat{\mathbf{F}}(k-1) +$
12: $\mu_0 \mathcal{F}\left\{\left[\left[\mathcal{F}^{-1}\left\{\mathbf{G}(k)\circ\mathbf{\Phi}(k)\right\}\right]_{1:N}^T \quad \mathbf{0}_N\right]^T\right\}$
13: **end for**
14: In our simulation we used the following set of parameters: $\mu_0 = 0.0325$, $\lambda_0 = 0.95$, $\lambda_1 = 0.9$.

Algorithm 3 Practical VSS [16] using FDAF-PEM-AFROW (PVSS-FDAF-PEM-AFROW)

1: Initialize: $\boldsymbol{S}_{U_a} = \boldsymbol{S}_{E_a} = \mathbf{0}_{M\times 1}$.
2: **for** $k = 1, 2, ...$ **do**
3: Perform lines $4-14$ from Algorithm 1
4: **for** $m = 0, 1, ..., M-1$ **do**
5: $S_{E_a}(m,k) = \lambda_0 S_{E_a}(m,k-1) + (1-\lambda_0)|E_a(m,k)|^2$
6: $S_{U_a}(m,k) = \lambda_0 S_{U_a}(m,k-1) + (1-\lambda_0)|U_a(m,k)|^2$
7: $\mu_{\text{PVSS}}(m,k) = \mu_0\left|1 - \dfrac{\sqrt{\hat{\sigma}_w(k)}}{\sqrt{S_{E_a}(m,k)+\alpha}}\right|$
8: $G(m,k) = \left(S_{U_a}(m,k) + \alpha\right)^{-1}$
9: **end for**
10: $\hat{\mathbf{F}}(k) = \hat{\mathbf{F}}(k-1) +$
11: $\mathcal{F}\left\{\left[\left[\mathcal{F}^{-1}\left\{\boldsymbol{\mu}_{\text{PVSS}}(k)\circ\mathbf{G}(k)\circ{\mathbf{U}_a}^*(k)\circ\mathbf{E}_a(k)\right\}\right]_{1:N}^T \quad \mathbf{0}_N\right]^T\right\}$
12: **end for**
13: In our simulations we used the following set of parameters: $\lambda_0 = 0.99$, $\alpha = 2e^{-2}$, $\mu_0 = 0.025$, $\alpha_0 = 10^{-8}$.

Algorithm 4 Projection-correlation VSS [11] using FDAF-PEM-AFROW (PCVSS-FDAF-PEM-AFROW)

1: Initialize: $\mu_{\text{PCVSS}}(0) = \mathbf{I}_{M\times 1}$, $\mathbf{S}_{U_a} = \mathbf{\Phi} = \mathbf{C} = \mathbf{0}_{M\times 1}$.
2: **for** $k = 1, 2, ...$ **do**
3: Perform lines $4-14$ from Algorithm 1
4: $\mathbf{\Phi}(k) = {\mathbf{U}_a}^*(k) \circ \mathbf{E}_a(k)$
5: **for** $m = 0, 1, ..., M-1$ **do**
6: $S_{U_a}(m,k) = \lambda_0 S_{U_a}(m,k-1) + (1-\lambda_0)|U_a(m,k)|^2$
7: $C(m,k) = C(m,k-1) + \Phi(m,k) - \Phi(m,k-B)$
8: $\mu'(m,k) = \lambda_1\ \mu_{\text{PCVSS}}(m,k-1) + \beta_0\ |C(m,k)|$
9: $\mu_{\text{PCVSS}}(m,k) = \begin{cases} 0, & \mu'(m,k) < 0 \\ \mu_0, & \mu'(m,k) > \mu_0 \\ \mu'(m,k), & \text{otherwise} \end{cases}$
10: $G(m,k) = \left(S_{U_a}(m,k) + \alpha_0\right)^{-1}$
11: **end for**
12: $\hat{\mathbf{F}}(k) = \hat{\mathbf{F}}(k-1) +$
13: $\mathcal{F}\left\{\left[\left[\mathcal{F}^{-1}\left\{\boldsymbol{\mu}_{\text{PCVSS}}(k)\circ\mathbf{G}(k)\circ{\mathbf{U}_a}^*(k)\circ\mathbf{E}_a(k)\right\}\right]_{1:N}^T \quad \mathbf{0}_N\right]^T\right\}$
14: **end for**
15: In our simulation we used the following set of parameters: $\lambda_0 = 0.95$, $B = 50$, $\beta_0 = \lambda_1 = 0.9$, $\alpha_0 = 1e^{-3}$, $\mu_0 = 0.03$.

APPENDIX A
INSTANTANEOUS PSEUDO-CORRELATION CALCULATION

The following recursions show how the signal $z(t)$ for the IPC measure is calculated in different adaptive filtering algorithms, namely, in NLMS, PEM-AFROW, FDAF-NLMS and FDAF-PEM-AFROW.

We first consider the NLMS update equation given as

$$e(t) = y(t) - \hat{\mathbf{f}}(t-1)\mathbf{u}(t) \tag{31}$$

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \mu\frac{e(t)}{\alpha + \mathbf{u}^T(t)\mathbf{u}(t)}\mathbf{u}(t). \tag{32}$$

A derivation similar to (24) and (26) then leads to

$$\begin{aligned} e_v(t) &= e(t) - e_x(t) \\ &= v(t) - \mu\sum_{i=1}^{t-1}\frac{e_v(i)}{\alpha + \mathbf{u}^T(i)\mathbf{u}(i)}\left(\mathbf{u}^T(i)\mathbf{u}(t)\right) \end{aligned} \tag{33}$$

and

$$z(t) = \frac{\mu}{N}\sum_{i=1}^{t-1}\frac{e_v(i)}{\alpha + \mathbf{u}^T(i)\mathbf{u}(i)}\left(\mathbf{u}^T(i)\mathbf{u}(t)\right) \tag{34}$$

$$= \frac{1}{N}\hat{\mathbf{f}}^T(t-1)\mathbf{u}(t) \tag{35}$$

with $\alpha = 1$. The recursion to compute $z(t)$ in PEM-AFROW can be derived similarly where the NLMS is then a special case with $\hat{\mathbf{a}} = [1 \quad \mathbf{0}_{n_A}] \quad \forall t$

$$\hat{\mathbf{a}} = \text{Levinson-Durbin}\left\{\mathbf{v}(t), n_A\right\} \tag{36}$$

$$\text{Computed every P samples.} \tag{37}$$

$$v_a(t) = \hat{\mathbf{a}}^T\bar{\mathbf{v}}(t) \tag{38}$$

$$u_a(t) = \hat{\mathbf{a}}^T\bar{\mathbf{u}}(t) \tag{39}$$

$$e_{av}(t) = v_a(t) - \hat{\mathbf{f}}^T(t-1)\mathbf{u}_a(t) \tag{40}$$

$$g(t) = \frac{1}{\mathbf{u}_a^T(t)\mathbf{u}_a(t) + \alpha} \tag{41}$$

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + g(t)\mathbf{u}_a(t)e_{av}(t) \tag{42}$$

$$z(t) = \frac{1}{N}\hat{\mathbf{f}}^T(t-1)\mathbf{u}_a(t), \tag{43}$$

Algorithm	Computation	Total
PVSS-APA	$2K^2N + 4KN + 15K + 8$	48116
PVSS-FDAF-PEM-AFROW	$\frac{21M\log_2 M + 34M}{N} + \frac{6M\log_2 M + 2M + n_A^2 + (4+4M)n_A}{N}$	672
PCVSS-APA	$2K^2N + 4KN + 4K + 4N + 12$	52080
PCVSS-FDAF-PEM-AFROW	$\frac{15M\log_2 M + 19M + BM}{N} + \frac{6M\log_2 M + 2M + n_A^2 + (4+4M)n_A}{N}$	520
VR-APA	$2K^2N + 4KN + 4K$	48086
VR-FDAF-PEM-AFROW	$\frac{15M\log_2 M + 31M}{N} + \frac{6M\log_2 M + 2M + n_A^2 + (4+4M)n_A}{N}$	534

TABLE IV
COMPLEXITY COMPARISON BY THE NUMBER OF FLOPS PER RECURSION. ONE FFT/IFFT IS $3M\log_2 M$ FLOPS, ONE $\sqrt{\cdot}$ IS M FLOPS AND $K = 4$.

where, in this case $\mathbf{v}(t) = [v(t), ..., v(t-P+1)]^T$, $\bar{\mathbf{v}}(t) = [v(t), ..., v(t-n_A+1)]^T$, $\bar{\mathbf{u}}(t) = [u(t), ..., u(t-n_A+1)]^T$ and $\mathbf{u}_a(t) = [u_a(t), ..., u_a(t-N+1)]$.

The recursion to compute $z(t)$ in FDAF-PEM-AFROW can be derived similarly where the FDAF-NLMS is then a special case with $\hat{\mathbf{a}} = [1 \quad \mathbf{0}_{n_A}] \quad \forall t$ with $\lambda = 0.99$ is written as,

$$\hat{\mathbf{a}}(k) = \text{Levinson-Durbin}\left\{[\mathbf{v}(k)]_{1:P}, n_A\right\} \tag{44}$$

$$u_a(m,k) = [u(kN+1+m), ..., u(kN+1+m-n_A)]\hat{\mathbf{a}}(k) \tag{45}$$

$$m = 1, ..., M \tag{46}$$

$$v_a(m,k) = [v(kN+1+m), ..., v(kN+1+m-n_A)]\hat{\mathbf{a}}(k) \tag{47}$$

$$m = 1, ..., M \tag{48}$$

$$\mathbf{U}_a(k) = \frac{1}{\sqrt{M}}\mathcal{F}\left\{\mathbf{u}_a^T(k)\right\} \tag{49}$$

$$\hat{\mathbf{e}}_a(k) = \left[\mathbf{v}_a(k) - \mathcal{F}^{-1}\left\{\hat{\mathbf{F}}(k-1) \circ \mathbf{U}_a(k)\right\}\sqrt{M}\right]_{N+1:M} \tag{50}$$

$$\mathbf{e}_{\mathbf{v}a}(k) = \frac{1}{\sqrt{M}}\mathcal{F}\left\{[\mathbf{0}_N \quad \hat{\mathbf{e}}_a(k)]\right\} \tag{51}$$

$$S_{U_a}(m,k) = \lambda S_{U_a}(m,k-1) + (1-\lambda)|U_a(m,k)|^2 \tag{52}$$

$$m = 1, ..., M \tag{53}$$

$$G(m,k) = \left(S_{U_a}(m,k) + \alpha\right)^{-1} \quad m = 1, ..., M \tag{54}$$

$$\hat{\mathbf{F}}(k) = \hat{\mathbf{F}}_a(k-1) \tag{55}$$

$$+ \mathcal{F}\left\{[\mathbf{I}_N \quad \mathbf{0}_N]^T \mathcal{F}^{-1}\left\{\mathbf{G}(k) \circ \hat{\mathbf{E}}_a(k) \circ \mathbf{U}_a^*(k)\right\}\right\} \tag{56}$$

$$\mathbf{z}(k) = \frac{1}{N}\left[\mathcal{F}^{-1}\left\{\hat{\mathbf{F}}(k-1) \circ \mathbf{U}_a(k)\right\}\sqrt{M}\right]_{N+1:M} \tag{57}$$

Here $(\cdot)^*$ represents a complex conjugation operation and $\mathbf{v}(k) = [v(kN-N+1), ..., v(kN+N)]^T$. It should be noted that the IPC measure is calculated using the signal $v_a(t)$ instead of $v(t)$, as given in (27).

REFERENCES

[1] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*. Berlin: Springer-Verlag, 2001.

[2] T. van Waterschoot, G. Rombouts, P. Verhoeve, and M. Moonen, "Double-talk-robust prediction error identification algorithms for acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 846–858, Mar. 2007.

[3] S. Haykin, *Adaptive Filter Theory*. Upper Saddle River, NY: Prentice Hall, 2002.

[4] P. S. R. Diniz, *Adaptive filtering: Algorithms and Practical Implementations*. Boston, MA: Springer, 2008.

[5] B. Widrow and M. Hoff, "Adaptive switching circuits," in *Proc. WESCON Conv. Rec.*, vol. 4, 1960, pp. 96–140.

[6] V. J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Process.*, vol. 41, no. 6, pp. 2075–2087, June 1993.

[7] Y. Zhang, J. A. Chambers, W. Wang, P. Kendrick, and T. J. Cox, "A new variable step-size LMS algorithm with robustness to nonstationary noise," in *Proc. 2007 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'07)*, vol. 3, Honolulu, Hawaii, USA, Apr. 2007, pp. 1349–1352.

[8] W.-P. Ang and B. Farhang-Boroujeny, "A new class of gradient adaptive step-size LMS algorithms," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 805–810, Apr. 2001.

[9] H.-C. Shin, A. H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 132–135, Feb. 2004.

[10] Y. Zhang, N. Li, J. A. Chambers, and Y. Hao, "New gradient-based variable step size LMS algorithms," *EURASIP J. Advances Signal Process.*, vol. 2008, no. 105, Article ID 529480, 9 pages, Jan. 2008.

[11] T. Creasy and T. Aboulnasr, "A projection-correlation algorithm for acoustic echo cancellation in the presence of double talk," in *Proc. 2000 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'00)*, vol. 1, Istanbul, Turkey, June 2000, pp. 436–439.

[12] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electronics and Communication in Japan*, vol. 67, no. 5, pp. 19–27, Aug. 1984.

[13] J. Benesty, H. Rey, L. R. Vega, and S. Tressens, "A nonparametric VSS NLMS algorithm," *IEEE Signal Process. Lett.*, vol. 13, no. 10, pp. 581–584, Oct. 2006.

[14] C. Paleologu, S. Ciochină, and J. Benesty, "Double-talk robust VSS-NLMS algorithm for under-modeling acoustic echo cancellation," in *Proc. 2008 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'08)*, Las Vegas, USA, Mar. 2008.

[15] M. A. Iqbal and S. L. Grant, "Novel variable step size NLMS algorithms for echo cancellation," in *Proc. 2008 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'08)*, Las Vegas, USA, Mar. 2008, pp. 241 – 244.

[16] C. Paleologu, J. Benesty, and S. Ciochină, "A variable step-size affine projection algorithm designed for acoustic echo cancellation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no. 8, pp. 1466 – 1478, Nov. 2008.

[17] H. Rey, L. R. Vega, S. Tressens, and J. Benesty, "Variable explicit regularization in affine projection algorithm: Robustness issues and optimal choice," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2096 – 2108, May 2007.

[18] Y.-S. Choi, H.-C. Shin, and W.-J. Song, "Adaptive regularization matrix for affine projection algorithm," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 12, pp. 1087–1091, Dec. 2007.

[19] W. Yin and A. S. Mehr, "A variable regularization method for affine projection algorithm," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 57, no. 6, pp. 476 – 480, June 2010.

[20] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory*. Upper Saddle River, New Jersey: Prentice Hall, 1993.

[21] T. van Waterschoot and M. Moonen, "Double-talk robust acoustic echo cancellation with continuous near-end activity," in *Proc. 13th European Signal Process. Conf. (EUSIPCO '05)*, Antalya, Turkey, Sep. 2005.

[22] L. Ljung, *System Identification: Theory for the user*. Englewood Cliffs, New Jersey: Prentice Hall, 1987.

[23] G. Rombouts, T. van Waterschoot, K. Struyve, and M. Moonen, "Acoustic feedback cancellation for long acoustic paths using a nonstationary source model," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3426–3434, Sep. 2006.

[24] T. Trump, "A frequency domain adaptive algorithm for colored measurement noise environment," in *Proc. 1998 IEEE Int. Conf.*

Acoust. Speech Signal Process. (ICASSP'98), vol. 3, Seattle, Washington, USA, May 1998, pp. 1705 – 1708.

[25] ——, "Accounting for measurement noise color in frequency domain adaptive algorithms," in *Proc. Thirty-Second Asilomar Conf. Signals Syst. Comp.*, vol. 1, Pacific Grove, CA, USA, May 1998, pp. 508 – 512.

[26] G. A. Clark, S. R. Parker and S. K. Mitra, "A Unified Approach to Time- and Frequency-Domain Realization of FIR Adaptive Digital Filters," *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 31, pp.1073-1083, Oct. 1983.

[27] T. Kariya, H. Kurata, *Generalised Least Squares*. John Wiley & Sons, 2004.

[28] T. Söderström and P. Stoica *System identification*, Prentice Hall, 1988.

[29] H. Buchner, J. Benesty and W. Kellermann, "Generalized multichannel frequency-domain adaptive filtering: efficient realization and application to hands-free speech communication," *Signal Process.*, vol. 85, pp. 549-570, Sep. 2005.

[30] P. Loizou, *Speech Enhancement: Theory and Practice*. Boca Raton, Florida: Taylor and Francis, 2007.

[31] R. M. Gray, "On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices," *IEEE Trans. on Information Theory*, vol.18, no.6, pp. 725-730, Nov. 1972.

[32] T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: state of the art and future challenges," *Proc. IEEE*, vol. 99, no. 2, pp. 288–327, Feb. 2011.

[33] J. M. Gil-Cacho, T. van Waterschoot, M. Moonen, and S. H. Jensen, "Transform domain prediction error method for improved acoustic echo and feedback cancellation," in *Proc. 20th European Signal Process. Int. Conf. (EUSIPCO'12)*, Bucharest, Rumania, Aug. 2012, pp. 2422–2426.

[34] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[35] V. Myllylä and G. Schmidt, "Pseudo-optimal regularization for affine projection algorithms," in *Proc. 2002 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'02)*, vol. 2, Orlando, Florida, May 2002, pp. 1917–1920.

[36] C. Rohrs and R. Younce, "Double talk detector for echo canceler and method," US Patent US 4 918 727, 04 17, 1990. [Online]. Available: http://www.patentlens.net/patentlens/patent/US4918727/

[37] J. M. Gil-Cacho, T. van Waterschoot, M. Moonen, and S. H. Jensen, "Wiener variable step size and gradient spectral variance smoothing for double-talk-robust acoustic echo cancellation and acoustic feedback cancellation," *Signal Process.*, vol. 104, pp. 1–14, Apr. 2014.

Jose Manuel Gil-Cacho was born in Salamanca, Spain, in 1980. He received the B.Sc. degree in electrical engineering in 2001-2005, from the Polytechnic University Madrid (UPM). He moved to England and received the M.Sc. degree in sound and vibration studies in 2007, from the Institute of Sound and Vibration Research (ISVR), Southampton University, UK. After the M.Sc. he was granted a one year Marie Curie scholarship to perform research in the aeroacoustic research group of LMS International, Belgium. Dr. Gil-Cacho finished his Ph.D. in 2013 at the at the electrical engineering Department of KU Leuven, Belgium. He is currently a consultant and his research interests are in new audio technologies and in adaptive signal processing with application to (nonlinear) acoustic signal enhancement, speech and audio processing

Toon van Waterschoot (S'04, M'12) received the MSc degree (2001) and the PhD degree (2009) in Electrical Engineering, both from KU Leuven, Belgium.

He is currently a part-time Assistant Professor at KU Leuven, Belgium, and a part-time Postdoctoral Research Fellow of the Research Foundation - Flanders (FWO), Belgium. He has previously held positions as a Teaching Assistant with the Antwerp Maritime Academy, Belgium (2002), as a Research Assistant with KU Leuven, Belgium (2002-2009) and with the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT), Belgium (2003-2007), and as a Postdoctoral Research Fellow at KU Leuven, Belgium (2009-2010) and at Delft University of Technology, The Netherlands (2010-2011). Since 2005, he has been a Visiting Lecturer at the Advanced Learning and Research Institute of the University of Lugano (Universit della Svizzera italiana), Switzerland. He has been teaching courses related to digital signal processing and control theory. His research interests are in signal processing, machine learning, and numerical optimization, with application to acoustic modeling, audio analysis, and acoustic signal enhancement.

Dr. van Waterschoot is serving as an Associate Editor for the Journal of the Audio Engineering Society and for the EURASIP Journal on Audio, Music, and Speech Processing, and as a Guest Editor for Signal Processing. He has been a Nominated Officer for the European Association for Signal Processing (EURASIP), and a Scientific Coordinator of the FP7-PEOPLE Marie Curie Initial Training Network on Dereverberation and Reverberation of Audio, Music, and Speech (DREAMS). He has been serving as an Area Chair for Speech Processing at the European Signal Processing Conference (EUSIPCO 2010, 2013, 2014, and 2015), and will be the General Chair of the 2016 AES Conference to be held in Leuven, Belgium. He is a member of the Audio Engineering Society, the Acoustical Society of America, EURASIP, and IEEE.

Marc Moonen (M'94, SM'06, F'07) is a Full Professor at the Electrical Engineering Department of KU Leuven, where he is heading a research team working in the area of numerical algorithms and signal processing for digital communications, wireless communications, DSL and audio signal processing. He received the 1994 KU Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), the 2004 Alcatel Bell (Belgium) Award (with Raphael Cendrillon), and was a 1997 Laureate of the Belgium Royal Academy of Science. He received a journal best paper award from the IEEE Transactions on Signal Processing (with Geert Leus) and from Elsevier Signal Processing (with Simon Doclo).

He was chairman of the IEEE Benelux Signal Processing Chapter (1998-2002), a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications, and President of EURASIP (European Association for Signal Processing, 2007-2008 and 2011-2012) He has served as Editor-in-Chief for the EURASIP Journal on Applied Signal Processing (2003-2005), and has been a member of the editorial board of IEEE Transactions on Circuits and Systems II, IEEE Signal Processing Magazine, Integration-the VLSI Journal, EURASIP Journal on Wireless Communications and Networking, and Signal Processing. He is currently a member of the editorial board of EURASIP Journal on Applied Signal Processing and Area Editor for Feature Articles in IEEE Signal Processing Magazine.

SHoldt Jensen was born in Haderslev, Denmark. He received the M.Sc. degree in electrical engineering from Aalborg University, Aalborg, Denmark in 1988, and the Ph.D. degree in signal processing from the Technical University of Denmark, Lyngby, Denmark in 1995.

Before joining the Department of Electronic Systems of Aalborg University, he was with the Telecommunications Laboratory of Telecom Denmark Ltd, Copenhagen, Denmark; the Electronics Institute of the Technical University of Denmark; the Scientific Computing Group of the Danish Computing Center for Research and Education (UNIC), Lyngby, Denmark; the Electrical Engineering Department of Katholieke Universiteit Leuven, Leuven, Belgium; and the Center for PersonKommunikation (CPK) of Aalborg University.

Dr. Jensen was an Associate Editor for the IEEE Transactions on Signal Processing, is Member of the Editorial Board of EURASIP Journal on Advances in Signal Processing, and is Associate Editor for the Research Letter in Signal Processing. He has also guest-edited two special issues for the EURASIP Journal on Applied Signal Processing on Anthropomorphic Processing of Audio and Speech and Digital Signal Processing in Hearing Aids and Cochlear Implants. He is a recipient of an European Community Marie Curie Fellowship, former Chairman of the IEEE Denmark Section, and Founder, Interim Chairman, and Chairman of the IEEE Denmark Section's Signal Processing Chapter. His research interests include statistical signal processing. speech and audio processing, multimedia technologies, digital communications, and satellite based navigation.

At Aalborg University he has been Member of the Study Board for Electronics and Information Technology and is currently Member of the Academic Council for the Faculty of Engineering, Science and Medicine. He is also PhD Study Director of the Doctoral School of Technology and Science at the Faculty of Engineering, Science and Medicine. In 2004 he received the Teacher of the Year Award in Electronics and Information Technology as voted by the student body.